ELSEVIER

Theoretical Computer Science 151 (1995) 29–36

Theoretical Computer Science

A topological property of rational ω-languages

André Arnold

LaBRI, Université Bordeaux I, 351 cours de la Libération, 33405 Talence Cedex, France

Abstract

We prove that any rational subset K of A^ω is the set of cluster points of A^ω equipped with some ultrametric distance having some specific properties.

1. Introduction

The set $\mathbb{N}^\omega$ of infinite sequences of natural numbers equipped with the Baire distance is a topological space. The first Borel classes of this space are: Π_1^0, the family of closed sets; Σ_2^0, the family of countable unions of closed sets; and Π_3^0, the family of countable intersections of Σ_2^0 sets.

In [1] the authors show that every Π_3^0 subset of $\mathbb{N}^\omega$ is the set of cluster points of $\mathbb{N}^\omega$ for a Π_1^0 ultrametric distance which refines the Baire distance.

Since a rational ω-language in A^ω is a Π_3^0 subset of $A^\omega \subseteq \mathbb{N}^\omega$, it is the set of cluster points of A^ω. Here we offer a direct proof of this result, which has also been directly proved by Priese and Nolte in [2] for a subclass of rational ω-languages, called 1-dimensional.

The paper is organized as follows. In Section 2 we recall some definitions and we precisely state the theorem we are to prove. In Section 3, we prove the theorem. For pedagogical reasons we prove it first in the case of closed rational languages, then in the case of deterministic languages, and finally in the general case.

2. Definitions

2.1. Words

Let us consider a finite alphabet A having at least two letters. Let A^* and A^ω be, respectively, the sets of finite and the set of infinite words on A. Let $A^\infty = A^* \cup A^\omega$ equipped with the prefix order denoted by $\leqslant$. For $u \in A^\infty$, $|u|$ is the length

SSDI 0304-3975(95)00046-1

of u ($|u| = \omega$ if $u \in A^\omega$), and for $n-1 < |u|$,

$$u(n) \text{ is the } n\text{th letter of } u, \text{ so that } u = \begin{cases} u(1)u(2)\cdots u(n) & \text{if } |u| = n \in \mathbb{N}, \\ u(1)u(2)\cdots u(n)\cdots & \text{if } |u| = \omega. \end{cases}$$

$$u[n] \text{ is the prefix of } u \text{ of length } n\text{: } u[n] = \begin{cases} u(1)u(2)\cdots u(n) & \text{if } n > 0, \\ \varepsilon, \text{ the empty word} & \text{if } n = 0. \end{cases}$$

The set of all words of A^* of length greater than n will be denoted by $A^{>n}$.

We denote by $\mathrm{LF}(K)$ the set $\{u[n] \mid u \in K, n \in \mathbb{N}\}$ of left factors of K and, for $K \subseteq A^*$, by $\vec{E}(K)$ the set $\{u \in A^\omega \mid \mathrm{LF}(u) \cap K \text{ is infinite}\}$. The set $\vec{E}(K)$ is often called the *Eilenberg limit* of K.

2.2. Distances

For a distance d on A^ω, we denote by $B_d(u,n)$ the open set $\{v \in A^\omega \mid d(u,v) < 2^{-n}\}$. A distance d is *ultrametric* if $\forall u,v,w \in A^\omega, d(u,w) \leqslant \sup(d(u,v), d(v,w))$. It is well known that an ultrametric distance taking its value in $\{0\} \cup \{2^{-n} \mid n \geqslant 0\}$ can be characterized by the following property.

Proposition 1. *A mapping $d : A^\omega \times A^\omega \to \{0\} \cup \{2^{-n} \mid n \geqslant 0\}$ such that $\forall u, v,\ d(u,v) = 0 \Leftrightarrow u = v$ is an ultrametric distance iff $\forall u,v \in A^\omega, \forall n \in \mathbb{N}, v \in B_d(u,n) \Rightarrow B_d(u,n) \subseteq B_d(v,n)$.*

A subset K of A^ω is *closed* for d if any u such that for any $n \in \mathbb{N}$ there is $u_n \in K$ satisfying $d(u,u_n) < 2^{-n}$ belongs to K.

The *adherence* of a subset K for d is the set

$$\mathrm{Adh}_d(K) = \{u \in A^\omega \mid \forall n \in \mathbb{N}, \exists v \in K : d(u,v) < 2^{-n}\}.$$

The following result is immediate.

Proposition 2. *A set K is closed for d iff $K = \mathrm{Adh}_d(K)$. For any set K, $\mathrm{Adh}_d(K)$ is closed for d.*

A point u is *isolated* for d if there exists an n such that $\forall v, d(u,v) < 2^{-n} \Rightarrow u = v$, or in other words, $\exists n : B_d(u,n) = \{u\}$. A point that is not isolated is a *cluster point*; it has the property: $\forall n, \exists u_n \neq u : d(u,u_n) < 2^{-n}$. The set of cluster points of A^ω for d will be denoted by $\mathscr{C}_d$.

The set of cluster points can be also defined by the following property.

Proposition 3. *Let U_n be the set $\{u \in A^\omega \mid B_d(u,n) \neq \{u\}\}$. Let C_n be $\bigcup_{u \in U_n} B_d(u,n)$. Then $\mathscr{C}_d = \bigcap_{n \geqslant 0} C_n$.*

2.3. Baire distance

The *Baire distance* on A^{ω} is the ultrametric distance δ defined by

$$\delta(u,v) = \begin{cases} 0 & \text{if } u = v, \\ 2^{-n}, \text{ with } n = \sup\{i \mid u[i] = v[i]\} & \text{otherwise.} \end{cases}$$

Therefore, $B_{\delta}(u,n) = u[n+1]A^{\omega}$.

The following result is immediate.

Proposition 4. *For any* $K \subseteq A^{\omega}$, $\mathrm{Adh}_{\delta}(K) = \vec{E}(LF(K))$.

We say that a distance d refines the Baire distance if $\forall u,v, d(u,v) \geqslant \delta(u,v)$, which is equivalent to: $\forall u, \forall n, B_d(u,n) \subseteq u[n+1]A^{\omega}$.

2.4. The theorem

The theorem we are to prove can be stated as follows.

For every rational language $K \subseteq A^{\omega}$, *there is an ultrametric distance* d *with the following properties:*

(i) K *is the set of cluster points for* d,

(ii) d *refines* δ,

(iii) $\forall n \in \mathbb{N}, \forall u \in A^{\omega}$, *the ball* $B_d(u,n)$ *(open for* d*) is closed for* δ,

(iv) $\forall n \in \mathbb{N}, \forall u \in A^{\omega}$, *if* $B_d(u,n) \neq \{u\}$, *then* $B_d(u,n)$ *is a rational subset of* A^{ω},

(v) *There are a number* k *and rational languages* L_i $(i = 1,\dots,k)$ *of* A^*, *rational languages* K_i $(i = 1,\dots,k)$ *of* A^{ω}, *such that* $C_n = \bigcup_{i=1}^{k}(L_i \cap A^{>n})K_i$.

Let us remark that the condition (v) is sufficient to prove that $\mathscr{C}_d$ is rational: By Proposition 3, $\mathscr{C}_d = \bigcap_{n \geqslant 0} C_n$. By applying Büchi's theorem establishing equivalence between rational languages and models of formulas of the monadic second order theory of $\mathbb{N}$, one can show that there exists a formula $P(x)$ such that C_n is the set of models of $P(n)$, thus, $\mathscr{C}_d$ is the set of models of $\forall x P(x)$ and is rational.

The proof of this theorem will be given for three different cases (closed languages, deterministic languages, the general case) because the construction of d in the first two cases is simpler and presents its own interest. The three proofs follow the same pattern. After giving the definition of d we give an explicit definition of the sets $B_d(u,n)$. On this definition one can check properties (ii), (iii), and (iv). The distance d is shown to be ultrametric by using the characterization of Proposition 1. Then we show (v) and we apply Proposition 3 to show (i).

3. Proofs of the theorem

3.1. K is closed

In case K is closed for the Baire distance δ, δ could be a candidate for d. Unfortunately, all points of A^{ω} are cluster points for δ. Thus we define d such that $d = \delta$

on K and that "isolates" all points of the complement of K. Formally, we define d by

$$d(u,v) = \begin{cases} 0 & \text{if } u = v, \\ 2^{-n}, \text{ with } n = \sup\{i \mid u[i] = v[i] \in \mathrm{LF}(K)\} & \text{otherwise.} \end{cases}$$

Notice that $\{i \mid u[i] = v[i] \in \mathrm{LF}(K)\}$ is never empty since $\varepsilon \in \mathrm{LF}(K)$. We have

$$B_d(u,n) = \begin{cases} u[n+1]A^{\omega} & \text{if } u[n+1] \in \mathrm{LF}(K), \\ \{u\} & \text{otherwise.} \end{cases}$$

Properties (ii), (iii), and (iv) are obvious. Let us assume that $v \in B_d(u,n)$. Then $v = u[n+1]v'$, hence, $u[n+1] = v[n+1]$ and $B_d(u,n) = B_d(v,n)$.

The set U_n is equal to $\{u \mid u[n+1] \in \mathrm{LF}(K)\}$. Thus $C_n = (\mathrm{LF}(K) \cap A^{>n})A^{\omega}$ and (v) is satisfied. Finally, $u \in \bigcap C_n$ iff u has arbitrary long prefixes in $\mathrm{LF}(K)$, i.e. $u \in \vec{E}(\mathrm{LF}(K)) = \mathrm{Adh}_{\delta}(K) = K$.

3.2. *K is deterministic*

A deterministic language $K \subseteq A^{\omega}$ is a language recognized by a deterministic Büchi automaton. Actually, we use in this paper the following equivalent definition (see Remark 4.1 of [4]).

A langage K is deterministic iff it is equal to $\vec{E}(L)$ for some rational language L of A^*. We define d by

$$d(u,v) = \begin{cases} 0 & \text{if } u = v, \\ 2^{-n}, \text{ with } n = \sup\{i \mid u[i] = v[i] \in L\} & \text{otherwise.} \end{cases}$$

Let us note that, when K is closed, $K = \vec{E}(\mathrm{LF}(K))$, thus the previous definition of d is a special case of this one. To be sure that $\{i \mid u[i] = v[i] \in L\}$ is never empty, we assume that $\varepsilon \in L$: adding ε to L does not change $\vec{E}(L)$.

For $u \in A^{\omega}$, let $L_n(u) \subseteq \mathrm{LF}(u)$ be the set $L \cap A^{>n} \cap \mathrm{LF}(u)$. We have

$$B_d(u,n) = \begin{cases} L_n(u)A^{\omega} & \text{if } L_n(u) \neq \emptyset, \\ \{u\} & \text{otherwise.} \end{cases}$$

Property (ii) is obvious. To show (iii) and (iv), we take z to be the shortest word in $L_n(u)$ when $L_n(u) \neq \emptyset$. Then $B_d(u,n) = L_n(u)A^{\omega} = zA^{\omega}$ because $z \in L_n(u) \subseteq zA^*$.

Let v be in $B_d(u,n)$. If $L_n(u)$ is empty then $v = u$, and $B_d(u,n) = B_d(v,n)$. If $L_n(u)$ is not empty then $v = zv'$, thus, $z \in L \cap A^{>n} \cap \mathrm{LF}(u) \cap \mathrm{LF}(v) \subseteq L_n(v)$. Hence, $B_d(u,n) = zA^{\omega} \subseteq L_n(v)A^{\omega} = B_d(v,n)$.

The set U_n is equal to $\{u \mid L_n(u) \neq \emptyset\}$. Thus $C_n = (L \cap A^{>n})A^{\omega}$ and (v) is satisfied. Finally, $u \in \bigcap C_n$ iff u has arbitrary long prefixes in L, i.e. $u \in \vec{E}(L) = K$.

3.3. *The general case*

Let $K \subseteq A^{\omega}$ be recognized by a deterministic Muller automaton $\mathscr{A} = \langle Q, \Delta, q_*, \mathscr{F} \rangle$ where Q is a set of states, Δ a mapping from $Q \times A$ to Q, $q_* \in Q$ the initial state, and

$\mathscr{F}$ a subset of $\wp(Q)$. The *q-run* of $\mathscr{A}$ on a word $u \in A^* \cup A^\omega$ is the unique sequence $\rho_q(u) = q_0 q_1 \ldots q_n, \ldots,$ or $q_0 q_1 \ldots q_{|u|}$ such that $q_0 = q$, $q_{i+1} = \Delta(q_i, u(i+1))$. We denote by $\gamma_q(u)$ the set of states occurring in $\rho_q(u)$. For $u \in A^\omega$, let $\mathrm{Inf}_q(u)$ be the set of states occurring infinitely often in the run $\rho_q(u)$. The word u is accepted by $\mathscr{A}$ if $\mathrm{Inf}_{q_*}(u) \in \mathscr{F}$.

Let us define the following rational languages:

- for any state q, L_q is the set of finite words such that q is the last state of the run $\rho_{q_*}(u)$,
- for $q, q' \in Q$ and $F \in \mathscr{F}$, $L^F_{q,q'}$ is the set of finite words such that

  q' is the last state of the run $\rho_q(u)$,

  $\gamma_q(u) = F$,

- for $q' \in Q$ and $F \in \mathscr{F}$, $L^{\subseteq F}_{q'}$ is the set of infinite words such that $\gamma_{q'}(u) \subseteq F$.

It can be easily checked that all the languages $L^{\subseteq F}_{q'}$ are closed for δ. For u and v in A^ω, let us define a set $L(u,v)$ of finite words by $x \in L(u,v)$ iff there is a $y \in A^*$ such that

– xy is a common prefix of u and v, i.e. $u = xyu'$ and $v = xyv'$,
– there are $q, q' \in Q$, there is $F \in \mathscr{F}$ such that $x \in L_q$, $y \in L^F_{q,q'}$, $u' \in L^{\subseteq F}_{q'}$, $v' \in L^{\subseteq F}_{q'}$.

We set $u \sqcap v$ equal to ε if $L(u,v) = \emptyset$, the longest word in $L(u,v)$ otherwise. Then, we define d by

$$d(u,v) = \begin{cases} 0 & \text{if } u = v, \\ 2^{-|u \sqcap v|} & \text{otherwise.} \end{cases}$$

Let us give the definition of $B_d(u,n)$ for $n > 0$. Let $I_n(u)$ be the set of triples $\langle q, q', F \rangle$ such that $F \in \mathscr{F}$ and $u \in (L_q \cap A^{>n}) L^F_{q,q'} L^{\subseteq F}_{q'}$. For $\langle q, q', F \rangle \in I_n(u)$, let $\mathrm{LF}^{>n}_{\langle q,q',F \rangle}(u)$ be the nonempty set $\{xy \mid u = xyu', x \in L_q \cap A^{>n}, y \in L^F_{q,q'}, u' \in L^{\subseteq F}_{q'}\}$. Then

$$B_d(u,n) = \begin{cases} \bigcup_{\langle q,q',F \rangle \in I_n(u)} \mathrm{LF}^{>n}_{\langle q,q',F \rangle}(u) L^{\subseteq F}_{q'} & \text{if } I_n(u) \neq \emptyset \\ \{u\} & \text{otherwise.} \end{cases}$$

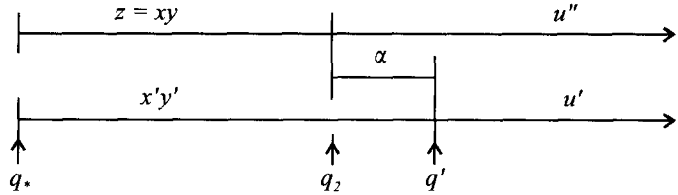
Property (ii) is still obvious. To show (iii) and (iv), we put $B_d(u,n)$ in another form. If $I_n(u) \neq \emptyset$, let z be the shortest word in $\bigcup_{\langle q,q',F \rangle \in I_n(u)} \mathrm{LF}^{>n}_{\langle q,q',F \rangle}(u)$, which is not empty. Let x be the shortest prefix of $z = xy$ such that $x \in L_q \cap A^{>n}$, $y \in L^F_{q,q'}$ for some $\langle q, q', F \rangle \in I_n(u)$ (such an x always exists) and let $\langle q_1, q_2, F_0 \rangle$ be the corresponding element of $I_n(u)$. Thus $z = xy$ and $u = xyu''$ with $|x| > n$, $x \in L_{q_1}$, $y \in L^{F_0}_{q_1,q_2}$, $u'' \in L^{\subseteq F_0}_{q_2}$. It follows

$$\gamma_{q_1}(y) = F_0, \qquad \gamma_{q_2}(u'') \subseteq F_0. \tag{1}$$

We claim that $B_d(u,n) = zL_{q_2}^{\subseteq F_0}$. Let $v \in B_d(u,n)$. Then $v = x'y'v'$, $u = x'y'u'$ with $|x'| > n$, $x' \in L_q$, $y' \in L_{q,q'}^F$, $u' \in L_{q'}^F$, $v' \in L_{q'}^F$, and $\langle q, q', F\rangle \in I_n(u)$. We have

$$\gamma_q(y') = F, \qquad \gamma_{q'}(u') \subseteq F, \qquad \gamma_{q'}(v') \subseteq F. \tag{2}$$

Moreover z is a prefix of $x'y'$. Let $x'y' = z\alpha$, thus $u'' = \alpha u'$.



We have to prove $v = x'y'v' = z\alpha v' \in zL_{q_2}^{\subseteq F_0}$, i.e. $\alpha v' \in L_{q_2}^{\subseteq F_0}$ or

$$\gamma_{q_2}(\alpha v') = \gamma_{q_2}(\alpha) \cup \gamma_{q'}(v') \subseteq F_0. \tag{3}$$

But $\gamma_{q_2}(\alpha) \subseteq \gamma_{q_2}(u'')$ and, by (1),

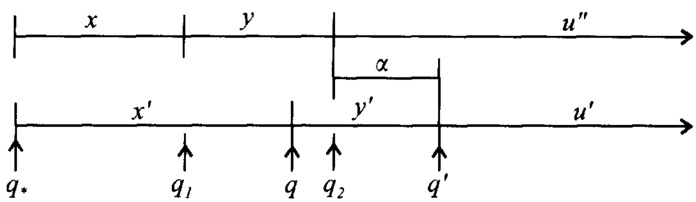
$$\gamma_{q_2}(\alpha) \subseteq F_0 = \gamma_{q_1}(y). \tag{4}$$

Thus (3) is implied by

$$\gamma_{q'}(v') \subseteq F_0 \tag{5}$$

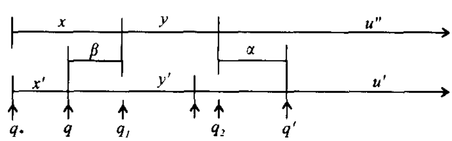
which is implied, because of (2), by $F \subseteq F_0$. Let us prove this inclusion.

We have two cases to consider: $|x| \leqslant |x'|$ and $|x| > |x'|$. In the first case:



we have $F = \gamma_q(y') \subseteq \gamma_{q_1}(y) \cup \gamma_{q_2}(\alpha) \subseteq F_0$, by (1)

In the second case:



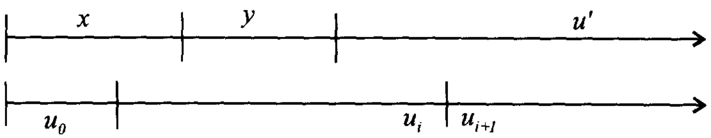
we have

$$\begin{aligned} F &= \gamma_q(\beta) \cup \gamma_{q_1}(y) \cup \gamma_{q_2}(\alpha) \\ &= \gamma_q(\beta) \cup \gamma_{q_1}(y) \quad \text{by (4)} \\ &= \gamma_q(\beta y). \end{aligned}$$

Thus $z = x'\beta y \in (L_q \cap A^{>n})L^F_{q,q_2}$. It follows that $\langle q, q_2, F\rangle \in I_n(u)$ since $u = x'\beta yu'$, hence $x' \in \bigcup_{\langle q,q',F\rangle \in I_n(u)} L_q \cap A^{>n}$, a contradiction with the fact that x is the shortest prefix of z having this propertry.

Let us show that d is an ultrametric distance. If $v \in B_d(u,n) = zL^{\subseteq F_0}_{q_2}$, then $\langle q_1, q_2, F_0\rangle \in I_n(v)$ and $zL^{\subseteq F_0}_{q_2} \subseteq B_d(v,n)$.

Now, let V_n be the set $\{u \in A^\omega \mid I_n(u) \neq \emptyset\}$, and let $D_n = \bigcup_{u \in V_n} B_d(u,n)$, for $n > 0$. We prove that $K = \bigcap_{n>0} D_n$. Let any word $u \in A^\omega$, and let $G = \mathrm{Inf}_{q_*}(u)$. The word u can be written $u_0u_1u_2 \ldots u_n \ldots$ with $u_i \in A^+$, and $u_0 \in L_{q_1}, u_{i+1} \in L^G_{q_{i+1},q_{i+2}}$ so that $u_0u_1 \ldots u_n \in L_{q_{n+1}}$, $u_{n+2}u_{n+3} \ldots \in L^{\subseteq G}_{q_{n+2}}$ and $u \in L_{q_{n+1}}L^G_{q_{n+1},q_{n+2}}L^{\subseteq G}_{q_{n+2}}$. If $u \in K$ then $G \in \mathscr{F}$ and $\langle q_{n+1}, q_{n+2}, G\rangle \in I_n(u)$, hence $u \in \bigcap_{n>0} D_n$. If $u \in \bigcap_{n>0} D_n$ then $u = xyu'$ with $|x| > |u_0|$, $x \in L_q$, $y \in L^F_{q,q'}$, $u' \in L^{\subseteq F}_{q'}$, and $F \in \mathscr{F}$. For i large enough, $|u_0u_1 \ldots u_i| > |xy|$.



Thus, $\gamma_{q_{i+1}}(u_{i+1}u_{i+2} \ldots) \subseteq \gamma_q(yu') \subseteq \gamma_{q_1}(u_1u_2 \ldots)$. But $\gamma_{q_{i+1}}(u_{i+1}u_{i+2} \ldots) = \gamma_{q_1}(u_1u_2 \ldots) = G, \gamma_q(yu') = F$. Hence, $F = G$ and $u \in K$.

It remains to prove that $D_n = C_n$, which is equivalent to $U_n = V_n$, i.e. $I_n(u) \neq \emptyset \Leftrightarrow B_d(u,n) \neq \{u\}$. By construction, $B_d(u,n) \neq \{u\} \Rightarrow I_n(u) \neq \emptyset$. But the converse is not true: a ball $B_d(u,n) = zL^{\subseteq F}_q$ may be a singleton if the closed rational language $L^{\subseteq F}_q$ contains only one word (which is ultimately periodic). To avoid this problem we slightly modify the previous construction so that $L^{\subseteq F}_q$ is never a singleton.

We add to A a new letter, say e, and we add to each state of the automaton recognizing K as a new transition labeled by e looping on this state. We get a new rational language K' with $K = K' \cap A^\omega$. Now, a language $L^{\subseteq F}_q$ is never a singleton: if a word v is obtained by inserting any number of e's in a word u of $L^{\subseteq F}_q$, then w is still in $L^{\subseteq F}_q$. Thus K' is the set of cluster points of $(A \cup \{e\})^\omega$. It remains to modify the distance d such that any word containing at least one occurrence of e is isolated. This is easily done by defining $d'(u,v) = \sup(d(u,v), 2^{-n})$ where n is the length of the longest common prefix of u and v which is in A^ω. It is easy to show that a point is a cluster point for d' iff it is a cluster point for d and it is in A^ω. Moreover, the balls $B_{d'}(u,n)$ are equal to $B_d(u,n) \cap A^{n+1}(A \cup \{e\})^\omega$ and $C'_n = C_nA^{n+1}(A \cup \{e\})^\omega$.

Acknowledgements

The definition of the distance d in the general case was obtained in collaboration with Serge Yoccoz. I thank the referee for pointing out an error in a previous definition of $B_d(u,n)$ in the general case.

References

[1] P. Darondeau, D. Nolte, L. Priese and S. Yoccoz, Fairness, distances and degrees, *Theoret. Comput. Sci.* **97** (1992) 131–142.
[2] D. Nolte and L. Priese, Strong fairness and ultra metrics, *Theoret. Comput. Sci.* **99** (1992) 121–140.
[3] L. Priese, Fairness, part II, *Bull. EATCS* **50** (1993) 247–259.
[4] W. Thomas, Automata on infinite objects, in: J. van Leeuwen, ed., *Handbook of Theoretical Computer Science, Vol. B* (Elsevier, Amsterdam, 1990) 133–191.